The Stata Journal (2004)
4, Number 1, pp. 89–92

Review of Veterinary Epidemiologic Research by Dohoo, Martin, and Stryhn

Laurent Audigé
AO Foundation
laurent.audige@aofoundation.org

Abstract. The new book by Dohoo, Martin, and Stryhn (2003) is reviewed.

Keywords: gn0010, veterinary, epidemiology

1 Introduction

Epidemiologists involved in research require a solid understanding of recognized methodology and statistical methods, their strengths and limitations, and their conditions of application. These methods have evolved tremendously over the years, becoming more powerful and also gaining in complexity. At the same time, the development of computer technology and the availability of powerful statistical software packages have made these techniques easier to apply, even by less experienced and expert researchers.

Such development is inevitable, and new standards in research are being set. Epidemiologists must learn to master and apply these standards, even in the context of a busy professional life. Many of us recognize that once a technique has been learned in a class or workshop, its application to real-life activities may not be straightforward without access to further professional guidance or the appropriate statistical software. This recently published book fills a gap in the epidemiological literature and provides researchers with a well-written, comprehensive presentation of current methods. The book has a substantial web site at *http://www.upei.ca/ver*, which includes a table of contents, a sample chapter, reviews and quotes, ordering information, feedback, datasets and programs, solutions to problems, and details on the authors.

2 Content and structure

Veterinary epidemiology has long been involved in the collection and analysis of data from observational studies (e.g., cross-sectional studies, cohort studies, case–control studies) in an attempt to explain the occurrence and development of animal health-related problems. In the introductory chapter, the authors report that many veterinary problems cannot be examined in the laboratory, and therefore, there is a need to study the "real world" in all its complexity. The main focus of this book is on the quantitative investigation and interpretation of "causal" relationships between study factors and health outcomes. Other areas in veterinary epidemiology, such as monitoring and surveillance systems, geographic information systems, and risk analysis, are not covered, but nevertheless, several chapters (e.g., sampling, measures of disease frequency, screen-

ing, and diagnostic tests) provide useful basic information. Basic methodological issues related to randomized controlled trials and meta-analysis are covered in two specific chapters.

Overall, this book is structured in two parts, following a logical path of research. The first part introduces basic concepts of disease causation, sampling, data-collection techniques (questionnaires, screening, and diagnostic tests), measures of associations, study designs, and validity issues. This part should allow researchers to define their study and develop appropriate data-collection tools. Confounding is addressed in a special chapter as it is a major issue when analyzing data from observational studies. The authors recommend the use of causal path diagrams showing how study factors relate to each other for an appropriate investigation of confounding bias. This research step can be completed during the planning phase of a study or, at the latest, during data collection (without knowledge of the data).

The rest of the book presents statistical techniques for the analysis of different data types from a very practical viewpoint. Essential formulas are provided for a solid understanding of major issues (assumptions, strengths, and limitations) in building statistical models and interpreting results, without being overwhelmed by complex notation. Specific chapters cover standard techniques, such as linear, logistic, multinomial, and Poisson regression, as well as survival analysis. Later chapters introduce the issue of data clustering and present more advanced techniques (e.g., mixed models for continuous and discrete data), taking into consideration the hierarchical structure of the epidemiological data. For instance, piglets may be grouped in litters, litters in sows, sows in farms, and farms in regions.

3 Software links and Stata material

I was impressed by the numerous excellent real-life examples, which reflect the wide experience and recognized expertise of the authors. Dataset examples can be obtained from the book's web site in Excel, SPSS, SAS, and Stata format. Most analyses were conducted using Stata 7, and commands were saved in do-files. As a Stata user, I tested all do-files provided on the book's web site with overall high satisfaction. I found it practical to run the analyses for each chapter after copying relevant files to a specific folder and replacing the location of files `c:\ver\data` by the current working directory. As I used Stata 8, I added `version 7` at the start of almost all do-files (this might be better added as a standard in downloadable files). At the time of this review, exercise results were becoming available on the book's web site. They are presented with the Stata command lines used, so readers can fully replicate the analyses.

To my knowledge, this is the first methodological book in this field that links the presentation of methods with practical applications in standard statistical software. This is an excellent initiative because meeting higher methodological standards in research will be facilitated by easy access to analytical tools. It would be worthwhile updating available Stata commands to the latest version of the software.

I agree with the authors about the danger of jumping too quickly into complex analyses and modeling without thorough prior examination of the data collected (especially if the tools are readily available). The benefit of following a structured, careful, and consistent approach to the analysis of "complex" observational databases cannot be stressed enough. Chapter 26 proposes a path of analysis that I would also highly recommend. For Stata users, it is easy to keep track of all analytical steps via do-files and to store the results in log files (following the NetCourse on Stata programming is very helpful in this respect).

4 Specific comments

While the book is very comprehensive and strong on techniques for observational studies, some comments may be made on a few specific issues.

The authors mention that screening and diagnostic tests may be assessed in validation studies before being used to collect data in epidemiological studies, but little information is provided on how to conduct such validation projects (although some elements are presented). It would be useful to include a structured paragraph on potential biases in the evaluation of diagnostic tests.

In addition, I feel the authors could have provided more practical guidelines in handling missing values, where researchers must decide between several options; e.g., impute the missing values, or drop the variables or the observations with missing values. The best option may depend on the absolute and relative frequencies of missing values in the exposure groups being compared.

Finally, an overview of meta-analysis is provided. Meta-analyses aim to summarize the results from a given series of comparative studies using a pooled-effect estimate. This is most commonly conducted for series of randomized controlled trials (RCTs) of treatment intervention, where confounding is of less concern than in observational studies and can also be used to summarize results from observational studies. The latter, however, should be done with care. Unless confounding bias has been well addressed by the authors of primary observational studies, reviewers may obtain spurious precision on an otherwise biased pooled-effect estimate (see the guidelines of the Non-Randomized Study Method Group of the Cochrane Collaboration). When considering the best available evidence about a "causal" association, conducting a systematic literature review is always preferred, but a meta-analysis may not be applied.

5 Conclusions

This book is about veterinary epidemiology, but I have no hesitation in recommending it to epidemiologists and researchers in human medicine. Many problems cannot be dealt with easily by conducting randomized controlled studies, e.g., for the study of interventions in populations, environmental factors, or surgery. In addition, complex data structures are found in human studies. Individuals are clustered in families, ethnic

groups, schools, villages, or cities, as well as regions and countries. Patients are also clustered in clinical practices and hospitals. Hence, methods to account for hierarchical data structures can be applied.

I tremendously enjoyed reading this book. I believe both experts and epidemiologists in training will find it very useful, and it will likely become a reference to be kept at hand for easy consultation. This is a concise summary of current analytical methods for observational studies. It will reassure researchers that they have done the right thing or help them to revise their approach to data analysis. I believe it will have a significant impact within the discipline of epidemiology and will contribute to an improvement in the quality of the research. The availability of Stata analysis do-files (or similar files for other statistical software packages, such as SAS) appropriate for most common research issues and techniques is very useful.

6 References

Dohoo, I., W. Martin, and H. Stryhn. 2003. *Veterinary Epidemiologic Research*. Charlottetown, PE, Canada: Atlantic Veterinary College.

About the Author

Laurent Audigé is a veterinary epidemiologist and manager of methodology at the AO Clinical Investigation and Documentation (AOCID), AO Foundation, in Dübendorf, Switzerland. His interests include the implementation and analysis of observational studies and randomized control trials, diagnostic issues related to health monitoring, and surveillance systems, as well as clinical decisions, meta-analyses in the context of systematic literature reviews, and validation studies for classification systems. He has been a Stata user since 1995.

The Stata Journal (2004)
4, Number 1, p. 93

Stata tip 4: Using display as an online calculator

Philip Ryan, University of Adelaide
philip.ryan@adelaide.edu.au

Do you use Stata for your data management, graphics, and statistical analysis but switch to a separate device for quick calculations? If so, you might consider the advantages of using Stata's built-in `display` command:

1. It is always at hand on your computer.
2. As with all Stata calculations, double precision is used.
3. You can specify the format of results.
4. It uses and reinforces your grasp of Stata's full set of built-in functions.
5. You can keep an audit trail of results and the operations that produced those results, as part of a log file. You can also add extra comments to the output.
6. Editing of complex expressions is easy, without having to re-enter lengthy expressions after a typo.
7. You can copy and paste results elsewhere whenever your platform supports that.
8. It is available via the menu interface (select **Data**—**Other utilities**—**Hand calculator**).
9. It can be abbreviated to `di`.

To be fair, there are some disadvantages, such as its lack of support for Reverse Polish Notation or complex number arithmetic, but in total, `display` provides you with a powerful but easy-to-use calculator.

```
. di _pi
3.1415927
. di %12.10f _pi
3.1415926536
. * probability of 2 heads in 6 tosses of a fair coin
. di comb(6,2) * 0.5^2 * 0.5^4
.234375
. di "chi-square (1 df) cutting off 5% in upper tail is " invchi2tail(1, .05)
chi-square (1 df) cutting off 5% in upper tail is 3.8414588
. * Euler-Mascheroni gamma
. di %12.10f -digamma(1)
0.5772156649
```

